Household Life Insurance Demand - a Multivariate Two-Part Model

Edward W. (Jed) Frees and Yunjie (Winnie) Sun

University of Wisconsin-Madison *

September 17, 2009

Abstract

What types of households own life insurance? To answer this question, we examine the Survey of Consumer Finances, a probability sample of the U.S. population. Household demand of two types of insurance, term and whole life insurance, is examined jointly. We model both the frequency and the severity of demand for insurance, building on the work of Lin and Grace (2007) by using explanatory variables that they developed. For the frequency portion, the household decisions about whether to own term and whole life insurance are modeled simultaneously with a bivariate probit regression model. Given ownership of life insurance by a household, the amounts of insurance are analyzed using generalized linear models with a normal copula for the severity portion. The copula permits the bivariate modeling of the insurance amount for households who own both term and whole life insurance, about 20% of our sample. Our findings suggest that household demand for term and whole life insurance is jointly determined. After controlling for explanatory variables, there exists a negative relationship for the frequency part and a positive relationship for the severity part. This mixed effect extends prior work which established a negative relationship, suggesting that term life insurance and whole life insurance are substitutes for one another. In contrast, our findings show that they are substitutes in the frequency yet complements in the severity.

*Keywords: Copulas, bivariate probit regression, generalized linear models.

1 Introduction

There is a large market for life insurance. For example, the direct premium written for life insurance in the U.S. reached about \$184 billion in year 2007 (Insurance Information Institute (2009)). Life insurance can be decomposed into two major categories, term life insurance which offers pure insurance protection, and whole life insurance which incorporates a savings component. In 2007, approximately 2.3 million term life insurance policies were sold by the top ten life insurers (by number of term life policies) with about \$442 billion insurance issued. Similarly, the top ten life insurers (by number of whole life policies) issued approximately 2.8 million whole life policies with about \$183 billion insurance written. Thus, life insurance is a significant segment of the insurance market. Understanding the characteristics of a household that drive life insurance demand can help insurers target their markets effectively and efficiently.

We use the Survey of Consumer Finances (SCF) to examine a household's demand for term and whole life insurance. In the insurance and consumer science literatures, much research has addressed how much life insurance protection households seek given their economic and demographic structure (see Goldsmith (1983), Burnett and Palmer (1984) and Lin and Grace (2007)). These literatures have focused primarily on middle-aged married couples and have provided solid economic justification underpinning life insurance demand. We therefore begin our analysis with middle-aged married couples using the explanatory variables suggested by the recent work of Lin and Grace (2007). The objective of our analysis is to explore the complicated relationship between the two types of life insurance using advanced analytic techniques.

In contrast to the traditional ordinary and censored (tobit) regression models that are widely applied in the life insurance demand literature, we propose a two-part model to analyze household demand for life insurance. In a two-part model, the frequency component (whether the household owns life insurance) and the severity component (how much life insurance is owned) are modeled separately (see Klugman et al. (2008)). Compared with ordinary regression which ignores the special pattern of a large portion of zeros in the dependent variable, a two-part model can provide an unbiased estimation. Though the tobit model takes the zeros into consideration, a drawback is that the same set of covariates are applied to both the magnitude as well as the censoring of the dependent variable. Cragg (1971) has showed several instances that the magnitude and the censoring of the dependent variable can be explained by different sets of covariates. Therefore, a model that can specify both the censoring (frequency) and the magnitude (severity) process is preferred. Many actuarial data sets come in two parts naturally; it is common to utilize a two-part model in such cases. See Frees (2010) for further discussion.

We study a household's demand for insurance in a multivariate framework, examining term life and whole life ownership in terms of frequency as well as severity components. Lin and Grace (2007) hypothesized that term life insurance is a substitute for whole life insurance and they incorporated the term life insurance face value linearly into the demand function of the whole life insurance. The method of Lin and Grace implies that household demand of whole life insurance depends on the demand of term life. In contrast, we treat these two demands as being jointly determined.

Through our joint model, we shed light on whether term and whole life insurance are complements or substitutes. Recall that two economic goods are substitutes if one can serve as a replacement for the other. They are complements if they "go together," that is, an increase in the demand for one good is aligned with an increase in the demand for the other good. About 20% of our sample are households who own both types of life insurance. It may be that unobserved information about households causes the complement/substitute relationship to differ between the frequency and severity components, a feature that our multivariate model allows. For example, consider an income-constrained, risk-adverse household. There may be a negative relationship in ownership due to income constraints yet a positive relationship in amounts due to conservative attitudes toward risk. Modeling the demand for these two types of insurances in a joint framework with all these considerations will provide more insight into the substitute or complement relationship of these two types of insurance.

This paper focuses on the analytical techniques to analyze household demand for life insurance. Section 2 describes the data. Section 3 describes the statistical models used and provides the empirical results of our analysis. Section 4 contains some concluding remarks.

2 Survey of Consumer Finance Data

2.1 Data Description

We use the 2004 Survey of Consumer Finances (SCF) data to conduct the analysis. The SCF is a triennial survey of U.S. families conducted by the Federal Reserve. The dataset is from a probability sample of the U.S. population with wealthy households over-sampled. The households participating in the survey were randomly selected to represent all economic strata of the country. Because of this, the sample survey results can be extrapolated to the national population. Unlike many insurance experience study datasets, we need not be concerned that our findings are limited by the underwriting or other practices of a limited group of insurers. The dataset includes extensive demographic and economic characteristics of the households as well as behavioral aspects such as "the motive to leave a bequest." The respondent has been designated to be the head of the household (see Federal Reserve (2004)).

The SCF data files imputed missing information by five methods for each household, providing five "implicates" data sets. Following Bernheim et al. (2001) and Lin and Grace (2007), we use the first implicate in this study.

The 2004 SCF data set has 4,492 household level ("primary economic unit") observations after we clean up some data errors. Among these households, 2,683 are married couples (or living with partner) while 921 are single-person households (never married, separated, divorced or widowed). About 77% of married couple households have some type of life insurance and for single person households, this rate is about 53%.

Following the insurance and consumer science literatures, we focus our analysis on married couples in the 20 to 64 age range. As in Lin and Grace (2007), we exclude those observations where labor income is paid by a lump sum, one payment only/in total, by the piece/job and whether the pay varies. In addition, we delete a few observations with missing or implausible values. The final number of observations used for the married couple analysis is $n = 2,150$.

There are two important limitations of the SCF data for our work. First, the SCF data are based on the whole household. Information about who the policyholders of the life insurance are and the face value for each policy is not available. This is an obstacle to understanding individual demand for life insurance (except for those single-person households). For married couples, we consider the whole household as a decision-making unit. It is appropriate to draw conclusions about the household's total life insurance demand based on the households' characteristics. However, when we make inference, we should be cautious of the generalization of the results.

Second, we do not know when the insurance was purchased. Compared with term life insurance, a short-term product, whole life insurance is a long-term commitment. The characteristics of the household may have been changed since the time that the whole life insurance was purchased. Using the current characteristics of the household to estimate a decision made in the past may introduce measurement error. Therefore, we define the demand for whole life insurance as the amount of whole life insurance that a household is willing to retain at the time of the survey.

2.2 Variables of interest

2.2.1 Dependent Variables

The dependent variables in our analysis measure two types of life insurance, term and whole life. For term life insurance, Table 1 shows that 1,416 of our 2,150 households, or 65.86%, own term life insurance. To measure the amount of demand, we examine the face amount of the policy (FACETerm). As is seen in Table 1 and Figure 1, the distribution is skewed to the right and heavy-tailed.

For the whole life component, Table 1 shows that 718 households, or 33.40%, purchased whole life insurance. For this amount, one could use the face value to measure the demand. Instead, we follow the insurance literature (e.g., Lin and Grace (2007)) and used the "net amount at risk" (NAR). The NAR is the difference between the face value and the saving component (cash value) of the whole life insurance. It can be interpreted as the amount of protection for unforeseen risk that a household has at the time of the survey.

Table 1: Life Insurance Summary Statistics

			Distribution of Positive Values				
Variable	Number	Percent	Minimum	25th Percentile	Median	75th Percentile	Maximum
FACETerm	1,416	65.86	0.8	100	270	1,000	150,000
NAR	718	33.40	0.66	60.25	202.5	900	45,000

Note: Monetary variables are in thousands.

To assess the relationship between term and whole life ownership in the frequency component, we look to joint ownership, that is, households who own both term and whole life insurance. There are 424 households, or 19.72%, who own both term and whole life insurance. If we assume the decisions of owning term life insurance and whole life insurance are independent, the probability of owning both types of life insurance should be 21.99% ($= 0.6586 \times 0.3340$). The empirical 19.72% sample probability of owning both types suggests a negative association. Of course, this calculation does not control for the effects of explanatory variables nor does it establish statistical significance.

To assess the relationship between term and whole life ownership in the amount component, we calculated the Spearman correlation between FACETerm and NAR for the 424 households who owned both types of life insurance. The correlation of 0.467 suggests that when a household decides to have both types of life insurance, the amount they choose to have is positively related, without controlling for any explanatory variables. This interesting finding encourages us to investigate the bivariate attributes of the data set.

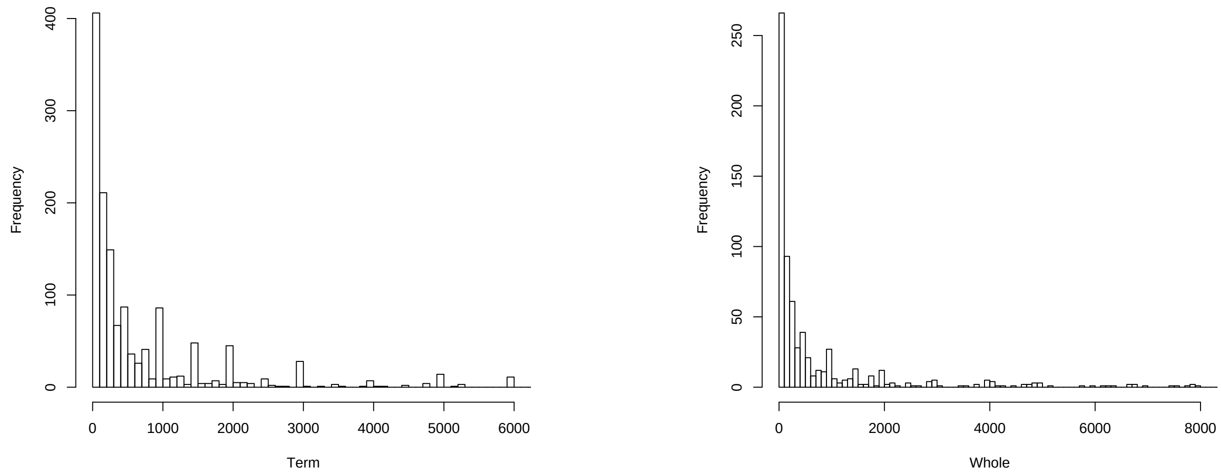

Figure 1: Histograms of Life Insurance Amounts (in thousands). The left-hand panel is for term insurance, the right-hand panel is for whole life. Both distributions are right-skewed and heavy-tailed.

2.2.2 Explanatory Variables

We build on the work of Lin and Grace (2007) by using explanatory variables they developed with some minor modifications. Their explanatory variables consist of the following categories:

- **Assets.** The role of assets in the demand for life insurance is unclear. One theory, based on the widely accepted decreasing absolute risk aversion (DARA), is that an increase in wealth will reduce the individual's willingness to insure (e.g., Chavas (2004)). Another position, increasing relative risk aversion (IRRA), suggests that individuals' risk aversion increases when they are subject to a given percent of wealth to risk. Many studies, including Halek and Eisenhauer (2001), have examined the existence of IRRA. The mixture of these two effects makes the relation between household assets and life insurance demand ambiguous.

  Lin and Grace (2007) split assets into eight categories: cash and cash equivalents, mutual funds, stocks, bonds, annuities, individual retirement accounts, real estate, and other assets. We follow such practice.

- **Debt.** The total debt of a household (DEBT) includes credit card, mortgage, line of credit, loan for home improvement, land contract loan, other real estate mortgage, car loan, education loan, consumer loans, other debt and margin loan. Life insurance on the one hand can protect the survivor from the financial burden of debt. On the other hand, debt might make insurance unaffordable for households under financial pressure. Therefore, the relationship between debt and life insurance demand is unclear.

- **Income.** The effect of the household income on the life insurance demand is similar to the wealth (asset) effect. Numerous studies (e.g., Goldsmith (1983)) have shown a positive relationship between these two. We hence expect that the higher the income, the greater are life insurance holdings. We use the regular salary and wage before taxes of the respondent (SALARY1) and the spouse (SALARY2) as measures of household income.

- **Bequests, Obligations, and Inheritance.** In the SCF data, there is a behavioral question about whether the household has a motive to leave a bequest. This is not only important for married couples but also important for single-person households. Vidal-Meliá and Lejárraga-García (2005) suggested that single people have a motivation to leave a bequest when they purchase life insurance or annuities. Some other questions that can influence the life insurance demand include the estimated amount of expected inheritance and the existence of foreseeable major financial obligation. To control for these factors are two binary variables indicating the desire to leave a bequest (BEQUEST) and the existence of financial obligation (OBLIGATION) and a continuous variable for the size of the inheritance expected (INHERITANCEexp).

- **Age.** In general, life insurance demand decreases as individuals age. This is because the accumulated wealth can reach the level that mostly meets the needs of the survivor. Another hypothesis is that since term life insurance is more affordable than whole life insurance, in the early stage of the life cycle when individuals earn relatively low income, they are more likely to purchase term life insurance than whole life insurance. From the convention employed in many similar investigations, we use the average age of the household to show the age effect and in addition, we add a quadratic term of the average age of the household to investigate the changing effect in age.

- **Education.** Education gives individuals opportunities to understand the importance of risk management especially through insurance purchases. Burnett and Palmer (1984) show that life insurance demand is positively related to education level of an individual. Another hypothesis is that if the spouse has higher human capital from better education, the household is less likely to purchase term life insurance on the husband. This has been tested by Goldsmith (1983). Due to the limitation of data, we can not test this hypothesis directly and the effect of education level of the household members on the overall life insurance demand is subject to examination. The number of years in school for the respondent (EDUCATION1) and the spouse (EDUCATION2) respectively measure the education level of the household.

- **Financial Vulnerability Index.** The financial vulnerability index (IMPACT), developed by Lin and Grace (2007) measures the adverse financial impact on a household in terms of living standard decline upon the death of one household member. It is a lifecycle measure that depends on the household's current living standard, the expected reduction in living standard, and the expected mortality. For a detailed calculations, see Appendix A. The hypothesis is that the higher the financial vulnerability index, the more life insurance is demanded.

  Compared with the Lin and Grace (2007) study, our IMPACT variable has about 5% extreme values that are far beyond the maximum reported in Lin and Grace (2007). One explanation is that Lin and Grace (2007) excluded about one percent of their data deemed to be outliers for their four-period analysis (1992, 1995, 1998, and 2001). Another possible explanation is that several households in the 2004 data are very wealthy. To improve the utility of this measure, we cap the value of IMPACT at 4, at its approximate 95th percentile. We add a binary variable (INDIMPACT) to indicate when the IMPACT value exceeds 4. The households with IMPACT greater than 4 (INDIMPACT=1) are usually those households with extremely high income on one person and low or zero income on the other person.

2.3 Summary Statistics

Table 2 summarizes the continuous explanatory variables and Table 3 provides means for the binary variables. Additional descriptions of the variables appear in Appendix B.

For our married couple sample, Table 2 shows that the median average age is 47.5, with 16 years of education for the respondent and 15 years for the spouse. The median salary was \$60,000 for the respondent and \$13,000 for the spouse. Due to the skewness nature of the covariates, we use a modified logarithm transformation (for example, LNBOND = log(1+ Bond)) to transform most of the explanatory variables (the modification was to handle zeros). The logarithmic transform is a natural choice given the log-link that we will introduce in Section 3.2.

Table 2 also emphasizes the preponderance of zeros for many of the asset and income variables. Because of this feature, we created binary variables to indicate the presence or absence of zero values in these categories. These binary variables appear in Table 3. For example, this table shows that 58.7% of households do not have any stock investments.

Compared with Lin and Grace (2007) who worked with 1992, 1995, 1998, and 2001 SCF data, households surveyed in 2004 tend to demand more life insurance. Our sample consists of couples who are older, better educated, having more children and wealth, and rely more on debt. The average financial vulnerability measured by the median is similar.

Table 2: Continuous Explanatory Variable Summary Statistics

Variable	Minimum	25th Percentile	Median	75th Percentile	Maximum
CASHEQV	0	3	17	98	32,628
FUND	0	0	0	20	57,500
STOCK	0	0	0	50	200,000
BOND	0	0	0	1	100,000
ANNUITY	0	0	0	0	200,000
RETIREMENT	0	0	52	272	35,000
REALESTATE	0	127	350	1,294	194,380
OTHASSETS	0	15	31	66	97,203
DEBT	0	13	110	286	121,686
SALARY1	0	29	60	163	80,112
SALARY2	0	0	13	40	2,700
INHERITANCEExp	0	0	0	0	906,060
AGE	21	39.5	47.5	54.5	64
EDUCATION1	1	12	16	17	17
EDUCATION2	0	12	15	16	17
IMPACT	0	0.049	0.113	0.340	1265.02

Note: Monetary variables are in thousands.

Table 3: Binary Explanatory Variable Means

Variable	INDCASHEQV	INDFUND	INDSTOCK	INDBOND	INDANNUITY
Mean	0.033	0.700	0.587	0.675	0.886
Variable	INDRETIREMENT	INDREALESTATE	INDOTHASSETS	INDDEBT	INDINHERITANCE
Mean	0.250	0.123	0.044	0.156	0.798
Variable	BEQUEST	OBLIGATION	INDIMPACT		
Mean	0.488	0.589	0.048		

3 Statistical Models

For notation, let r_{i1} and r_{i2} be binary variables that indicate whether household i purchases term life insurance and whole life insurance, respectively. Similarly, let y_{i1} and y_{i2} denote the amounts, if available. We decompose the joint distribution of the dependent variables into frequency and severity components by

$$\mathrm{f}(r_{i1}, r_{i2}, y_{i1}, y_{i2}) = \mathrm{f}_F(r_{i1}, r_{i2}) \times \mathrm{f}_S(y_{i1}, y_{i2}|r_{i1}, r_{i2}).$$

For the frequency component $\mathrm{f}_F(r_{i1}, r_{i2})$, we employ a bivariate probit regression model. For the conditional severity component $\mathrm{f}_S(y_{i1}, y_{i2}|r_{i1}, r_{i2})$, we use generalized linear models with a copula.

3.1 Frequency Models

A bivariate probit regression model extends the probit regression model to a two-dimensional vector of binary responses (e.g., Greene (2008)). This distribution can be easily computed using a standard bivariate normal distribution.

To assess the frequency distribution f_F, we note that there are four possible outcomes for the ith observation. These are:

$$\begin{aligned}
\Pr(r_{i1} = 1, r_{i2} = 1) &= \Phi_2(\mathbf{x}_i'\boldsymbol{\beta}_1, \mathbf{x}_i'\boldsymbol{\beta}_2; \rho) \\
\Pr(r_{i1} = 1, r_{i2} = 0) &= \Pr(r_{i1} = 1) - \Pr(r_{i1} = 1, r_{i2} = 1) = \Phi(\mathbf{x}_i'\boldsymbol{\beta}_1) - \Phi_2(\mathbf{x}_i'\boldsymbol{\beta}_1, \mathbf{x}_i'\boldsymbol{\beta}_2; \rho) \\
\Pr(r_{i1} = 0, r_{i2} = 1) &= \Phi(\mathbf{x}_i'\boldsymbol{\beta}_2) - \Phi_2(\mathbf{x}_i'\boldsymbol{\beta}_1, \mathbf{x}_i'\boldsymbol{\beta}_2; \rho) \\
\Pr(r_{i1} = 0, r_{i2} = 0) &= 1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta}_1) - \Phi(\mathbf{x}_i'\boldsymbol{\beta}_2) + \Phi_2(\mathbf{x}_i'\boldsymbol{\beta}_1, \mathbf{x}_i'\boldsymbol{\beta}_2; \rho),
\end{aligned}$$

where $\Phi_2(\cdot)$ is the cumulative distribution function of the standard bivariate normal distribution with correlation parameter ρ. With these expressions, the frequency log-likelihood of the ith observation is

$$\begin{aligned}
l_{Fi} &= r_{i1}r_{i2} \ln \Phi_2(\mathbf{x}_i'\boldsymbol{\beta}_1, \mathbf{x}_i'\boldsymbol{\beta}_2; \rho) \\
&\quad + r_{i1}(1 - r_{i2}) \ln[\Phi(\mathbf{x}_i'\boldsymbol{\beta}_1) - \Phi_2(\mathbf{x}_i'\boldsymbol{\beta}_1, \mathbf{x}_i'\boldsymbol{\beta}_2; \rho)] \\
&\quad + (1 - r_{i1})r_{i2} \ln[\Phi(\mathbf{x}_i'\boldsymbol{\beta}_2) - \Phi_2(\mathbf{x}_i'\boldsymbol{\beta}_1, \mathbf{x}_i'\boldsymbol{\beta}_2; \rho)] \\
&\quad + (1 - r_{i1})(1 - r_{i2}) \ln[1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta}_1) - \Phi(\mathbf{x}_i'\boldsymbol{\beta}_2) + \Phi_2(\mathbf{x}_i'\boldsymbol{\beta}_1, \mathbf{x}_i'\boldsymbol{\beta}_2; \rho)].
\end{aligned}$$

The total frequency likelihood is $\sum_{i=1}^{n} l_{Fi}$. With this, it is straight-forward to determine maximum likelihood estimators using readily available statistical software.

3.2 Severity Model

To understand the joint distribution of the insurance amounts, it is natural to consider logarithmic amounts and use a bivariate normal distribution. This we did as a robustness check; Section C.2 shows that this is a sensible approach.

However, so as not to be constrained by bivariate normality, this paper explores a copula framework for analyzing the joint distribution of life insurance amounts. See Nelson (1999), Frees and Valdez (1998) and Genest and Favre (2007) for introductions to copula modeling. One important advantage of this approach is that one is not constrained to using (log) normal marginal distributions, although this is certainly a possibility. Historically, one of the first studies of life insurance coverage amounts, by Norwegian actuary Birger Meidell in 1912, used a Pareto distribution, a long-tailed distribution (Kleiber and Kotz (2003), p. 62).

Another advantage of the copula approach is that "copulas preserve the marginals." That is, by specifying a copula to model the relationship, marginal distribution are maintained when only a single amount is examined. This is important for our data where 59.8% ($= 0.6586 + 0.3340 - 2 \times 0.1972$) own only one type of life insurance.

Using copulas for the generalized linear model has been widely applied in biomedical and financial risk management literatures. Our work is most closely related to Frees and Wang (2005) who used Gaussian and t copulas with marginal gamma regression models to fit aggregate automobile claims.

Beginning with the joint severity distribution function f_S, we let $\mathrm{f}_{S1}(y_{i1}) = \mathrm{f}_S(y_{i1}, y_{i2}|r_{i1} = 1, r_{i2} = 0)$ be the marginal distribution for term and similarly define $\mathrm{f}_{S2}(y_{i2})$ for whole life. This sequence of definitions is viable because copulas preserve the marginals. We assume that each distribution is parameterized by a location parameter μ_{ij} and scale parameter ϕ_j, $j = 1$ for term and $j = 2$ for whole life. Although not critical to the development, we use a generalized linear model (GLM) framework with a logarithmic link so that $\mu_{ij} = \exp(\mathbf{x}_{ij}'\boldsymbol{\beta}_j)$, $j = 1, 2$.

We use a copula to express the relationship between the marginal distributions. Thus, we can write the bivariate distribution function as

$$\begin{aligned} \mathrm{f}_S(y_{i1}, y_{i2}) &= \mathrm{f}_S(y_{i1}, y_{i2}|r_{i1} = 1, r_{i2} = 1) \\ &= \mathrm{f}_{S1}(y_{i1})\mathrm{f}_{S2}(y_{i2})\mathrm{c}(\mathrm{F}_{S1}(y_{i1}), \mathrm{F}_{S2}(y_{i2})), \end{aligned} \tag{1}$$

where $\mathrm{F}_{Sj}(y_{ij})$ is the distribution function corresponding to $\mathrm{f}_{Sj}(y_{ij})$, $j = 1, 2$. This function $\mathrm{c}(\cdot)$ is called a copula density function. Typically copulas are parameterized by one or two parameters that describe the dependence and other features. In our empirical work, we focus on the normal copula with correlation parameter ρ_S.

The log-likelihood of the ith household's life insurance demand given they own term and whole life insurance is

$$l_{S12i} = \ln \mathrm{f}_{S1}(y_{i1}) + \ln \mathrm{f}_{S2}(y_{i2}) + \ln \mathrm{c}(\mathrm{F}_{S1}(y_{i1}), \mathrm{F}_{S2}(y_{i2})).$$

Thus, the total severity log-likelihood for the ith household is

$$l_{Si} = r_{i1}(1 - r_{i2}) \ln \mathrm{f}_{S1}(y_{i1}) + (1 - r_{i1}) r_{i2} \ln \mathrm{f}_{S2}(y_{i2}) + r_{i1} r_{i2} \times l_{S12i}.$$

The total severity likelihood is $\sum_{i=1}^{n} l_{Si}$. With this, it is straight-forward to determine maximum likelihood estimators using readily available statistical software.

3.3 Results

As described earlier, although our statistical models differ we build on the work of Lin and Grace (2007) by using the explanatory variables that they developed. Our reported models feature two main differences in the selection of explanatory variables. First, as described in Section 2.2.2, we included binary variables to indicate zero values for income and related financial variables. The motivation is similar to the intuition of the two-part model for dependent variables; we supplement the explanatory variable for the amount with a binary variable that indicates its presence or absence. For example, we have a variable to indicate whether a household invests in stocks as well as a variable to indicate the amount of stock investment.

Second, Lin and Grace (2007) hypothesized that a household's life insurance demand depends on asset variables and that dependence varies by three age groups (20-34, 35-49, and 50-64). From a preliminary investigation, we did not see any significant results supporting this hypothesis. Instead of using the interaction terms between age groups and assets covariates, we use the average age of the household as well as the quadratic form of the age to control the nonlinear effects of age. Although not detailed here, the main results that we report on the associations are not qualitatively impacted by these choices of explanatory variables.

Besides the analysis for married couples, we also fit the model to single person households. Due to the nature of single person households, we can make inference about individual's life insurance demand. Appendix D displays the result.

3.3.1 Frequency Model

The bivariate probit regression for the frequency model has 63 coefficients. Table 4 shows the result when the dependent variable is a vector of two binary indicators for the term life insurance and whole life insurance. Of the assets variables, only stock, bond and real estate impact both the term and whole life insurance demand decision. In general, the more assets a household has, the less likely a household is to own life insurance. However, for the real estate variable, it is negatively related to the probability of term life insurance ownership although positively related to that of whole life. One explanation is that when the household has more real estate, they prefer life insurance with saving component to diversify their investment. Most indicator variables for zero assets have negative coefficients. This means when a household does not have some type of assets, they are less likely to own life insurance. It may be that life insurance is not affordable for them.

The amount of debt a household owes positively affects the decision to hold term life insurance. The more debt a household owes, the more likely that it is to hold term life insurance protection, which is relatively inexpensive. The income level of the household also positively relates to the decision to hold life insurance. This is consistent with previous studies such as Lin and Grace (2007). The desire to leave a bequest only has positive impact on the likelihood of demand of whole life insurance.

The quadratic term of the age variable has negative coefficient while the original age variable has positive coefficient on term life insurance demand. This implies that when the members of the household get older, they are more likely to demand life insurance. However, the probability increases in a decreasing manner. The education level of the head of the household positively relates to the decision to hold term life insurance. This confirms Burnett and Palmer (1984)'s finding.

Surprisingly, the financial vulnerability variable proposed by Lin and Grace (2007) only has impact on the frequency of term life insurance demand but not on the whole life insurance demand. The larger the financial vulnerability index, the more likely that a household demands term life insurance. However, when the financial vulnerability index is extremely large, those households are less likely to own term life insurance. In fact, a very large financial vulnerability index implies that a household has extreme high income on one person and low or zero income on the other.

The correlation between the likelihood of term life insurance demand and the likelihood of whole life insurance demand is significantly negative after controlling for the covariates. This indicates that term life insurance and whole life insurance are substitutes in frequency. The greater the probability for one type, the smaller is the probability of holding the other type of life insurance.

3.3.2 Severity model

For the severity model, the sample includes only those households that own at least one type of life insurance. There are 65 coefficients in this model, including the scale parameters for the GLM and the correlation parameter for the Gaussian copula. Table 5 shows the result.

Unlike the frequency model (as well as results in Lin and Grace (2007)), the coefficients of almost all asset variables are positive. This suggests that the more assets a household has, the more life insurance they demand when they decide to have life insurance. This is in some sense consistent with the IRRA hypothesis. However, the indicators for all zero asset variables surprisingly show highly positive coefficients as well. This may be the case that for those less wealthy households with zero assets, they need more life insurance protection when the breadwinner deceases.

The severity model coefficients of debt, age and squared age of the couple, education level of the household, and salary of the household head have the same signs in the frequency model. For the spouse salary, the coefficient is negative for the severity model, implying that the more salary the spouse earns, the less is the need for higher amounts of life insurance.

The inheritance expected has a positive impact on the amount of the insurance demand. This is inconsistent with the intuition. The desire to leave a bequest and the existence of foreseeable major obligation also drive the household to demand more life insurance.

The financial vulnerability index and the indicator for extreme index value are both statistically significant in the severity model. The higher the financial vulnerability index, the more life insurance protection a household seeks. However, for those households who have one extremely high income member (usually very high IMPACT value and therefore INDIMPACT=1), they demand less life insurance.

The main result is that the correlation is positive and significant. This is opposite to the result from the frequency model. This implies when a household decides to own both type of life insurance, the more they demand on one type, the more they demand on the other type. Therefore, term life insurance demand and whole life insurance demand are complements in severity.

As a robustness check, we also fit the data with t copula and marginal gamma regression model. The fitted degree of freedom for the t copula is 32.49, which implies that the fitted t copula is approximately normal. The result from this model is very comparable to the results with Gaussian copula, therefore not reported.

3.4 Tobit Model

Tobit regression is a well-known left-censored regression model first proposed by Tobin (1958). An unobserved, or latent, variable is assumed to follow a linear regression model. The responses are censored or "limited" in the sense that observations are bounded from below, typically by zero. One drawback of the tobit model is its reliance on the normality assumption of the latent response. A second, and more important, drawback is that a single latent variable dictates both the magnitude of the response as well as the censoring. For example, when studying healthcare expenditures, a zero represents a person's choice or decision not to utilize healthcare during a period. For many studies, the amount of healthcare expenditure is strongly influenced by a healthcare provider (such as a physician); the decision to utilize and the amount of healthcare can involve very different considerations. See for example Frees, Gao and Rosenberg (2009) for further discussion.

In our study of a model for term life insurance demand, the coefficient for the salary of the spouse is positive while in the severity model, this coefficient is negative. A tobit model can not catch such differences that can be important for understanding the problem. We also fit the tobit model to our data set to compare with our two-part model and Lin and Grace (2007)'s results. Table 6 displays the result.

The results from tobit model reveal that the relationship between these two types of life insurance is negative, which supports Lin and Grace (2007)'s finding that these two insurances are substitutes. However, in our two-part model, the frequency part shows a negative relationship between term and whole life insurance ownership, while the severity part shows a positive relationship after controlling all the explanatory variables. This inconsistent result implies that some important attributes of the data may be overlooked when using the tobit model.

Table 4: Bivariate Probit Regression

	Term Life			Whole Life		
Parameter	Estimate	t-statistic		Estimate	t-statistic	
Intercept	0.667	0.724		-0.939	-0.992	
Log (1+ cash and cash equivalent)	0.030	1.593		0.042	2.164	**
Indicator for zero cash and cash equivalent	-0.241	-1.036		0.290	1.069	
Log (1+ fund)	0.031	1.227		-0.044	-1.795	*
Indicator for zero fund	0.344	1.133		-0.697	-2.381	**
Log (1+stock)	-0.052	-2.544	**	-0.037	-1.855	
Indicator for zero stock	-0.425	-1.854	*	-0.477	-2.160	**
Log (1+ bond)	-0.040	-2.405	**	-0.037	-2.335	**
Indicator for zero bond	-0.440	-2.657	***	-0.547	-3.525	***
Log (1+ annuity)	-0.072	-1.853		0.023	0.620	
Indicator for zero annuity	-0.872	-1.788		0.049	0.107	
Log (1+ retirement)	0.024	1.072		-0.032	-1.433	
Indicator for zero retirement	-0.122	-0.481		-0.388	-1.523	
Log (1+ real estate)	-0.209	-5.336	***	0.090	2.257	**
Indicator for zero real estate	-2.581	-5.684	***	0.818	1.739	*
Log (1+ other assets)	0.038	1.384		0.011	0.421	
Indicator for zero other assets	0.372	1.179		-0.339	-1.014	
Log (1 + debt)	0.056	2.307	**	0.005	0.182	
Indicator for zero debt	0.195	0.656		-0.002	-0.006	
Log (1+ salary of the respondent)	0.018	2.280	**	0.004	0.490	
Log (1+ salary of the spouse)	0.014	2.323	**	0.015	2.443	**
Indicator for the desire to leave a bequest	-0.003	-0.042		0.114	1.681	*
Indicator for foreseeable major financial obligation	0.075	1.201		-0.001	-0.008	
Log (1+ sizable inheritance expected)	-0.023	-0.641		-0.011	-0.294	
Indicator for zero inheritance expected	-0.323	-0.687		-0.172	-0.368	
Average age of the couple	0.058	2.240	**	0.003	0.123	
Squared average age of the couple	-0.001	-2.105	**	0.000	0.670	
Education level of the resondent	0.058	3.470	***	-0.017	-0.985	
Education level of the spouse	0.021	1.387		0.014	0.866	
Financial Vulnerability Index (IMPACT)	0.170	2.672	***	0.056	0.969	
Indicator for IMPACT $\geqslant 4$	-0.473	-1.933	*	-0.162	-0.727	
Rho	-0.285	-7.668	***			

Notes: Number of observations: 2,150.

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level

Table 5: Gaussian Copula with Gamma Marginals

Parameter	Term Life Estimate	Term Life t-statistic		Whole Life Estimate	Whole Life t-statistic	
Intercept	0.669	0.903		0.130	0.118	
Log (1+ cash and cash equivalent)	0.171	8.545	***	0.024	0.855	
Indicator for Izero cash and cash equivalent	1.196	3.859	***	-1.115	-2.078	**
Log (1+ fund)	0.030	1.218		0.056	1.542	
Indicator for zero fund	0.396	1.356		0.935	2.156	**
Log (1+stock)	0.044	2.206	**	0.075	2.531	**
Indicator for zero stock	0.415	1.882	*	1.001	2.994	***
Log (1+ bond)	0.063	3.588	***	0.074	3.279	***
Indicator for zero bond	0.457	2.874	**	0.625	2.795	***
Log (1+ annuity)	0.016	0.458		0.067	1.176	
Indicator for zero annuity	0.257	0.623		0.628	0.887	
Log (1+ retirement)	0.023	1.080		0.091	2.858	***
Indicator for zero retirement	0.175	0.713		0.753	1.954	*
Log (1+ real estate)	0.201	5.779	***	0.326	5.428	***
Indicator for zero real estate	2.195	5.435	***	3.506	4.632	***
Log (1+ other assets)	0.174	5.939	***	0.196	4.957	***
Indicator for zero other assets	1.825	5.220	***	1.286	2.385	**
Log (1 + debt)	0.129	5.263	***	0.040	0.990	
Indicator for zero debt	1.054	3.386	***	0.868	1.673	*
Log (1+ salary of the respondent)	0.017	1.994	*	0.012	0.976	
Log (1+ salary of the spouse)	-0.024	-3.951	***	-0.028	-2.908	***
Indicator for the desire to leave a bequest	0.206	3.097	***	0.635	5.758	***
Indicator for foreseeable major financial obligation	0.087	1.391		0.163	1.710	*
Log (1+ sizable inheritance expected)	0.163	4.504	***	0.041	0.696	
Indicator for zero inheritance expected	1.963	4.261	***	0.563	0.745	
Average age of the couple	0.023	2.674	***	0.022	1.832	*
Squared average age of the couple	-0.001	-5.700	***	-0.001	-5.141	***
Education level of the resondent	0.046	2.604	**	0.006	0.203	
Education level of the spouse	0.024	1.349		0.056	2.074	**
Financial Vulnerability Index (IMPACT)	0.105	1.791	*	0.253	2.733	***
Indicator for IMPACT $\geqslant$	-0.464	-1.970	*	-0.815	-2.384	**
Scale	0.913	0.032	$	0.746	0.024	$
Rho	0.099	1.964	*			

Notes: Number of observations: 1,710.

$ This is the standard error for the scale parameter.

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level

Table 6: Tobit Models for Term and Whole Life

	Term Life			Whole Life		
Parameter	Estimate	t-statistic		Estimate	t-statistic	
Intercept	1.739	0.308		-1.485	-0.135	
Log (1+ cash and cash equivalent)	0.317	2.667		0.497	2.175	**
Indicator for zero cash and cash equivalent	-0.997	-0.646		1.025	0.305	
Log (1+ fund)	0.086	0.556		-0.231	-0.823	
Indicator for zero fund	0.513	0.276		-4.886	-1.443	
Log (1+stock)	-0.455	-3.604	***	-0.458	-1.991	*
Indicator for zero stock	-4.195	-3.032	***	-5.832	-2.268	**
Log (1+ bond)	-0.329	-3.211	***	-0.534	-2.906	***
Indicator for zero bond	-3.672	-3.780	***	-7.574	-4.197	***
Log (1+ annuity)	-0.327	-1.409		0.111	0.265	
Indicator for zero annuity	-4.209	-1.491		-1.210	-0.234	
Log (1+ retirement)	0.117	0.853		-0.158	-0.608	
Indicator for zero retirement	-1.188	-0.771		-3.104	-1.037	
Log (1+ real estate)	-0.695	-2.849	***	0.625	1.329	
Indicator for zero real estate	-9.125	-3.223	***	2.425	0.436	
Log (1+ other assets)	0.195	1.143		0.410	1.315	
Indicator for zero other assets	1.346	0.674		-1.019	-0.261	
Log (1 + debt)	0.382	2.481	**	0.130	0.431	
Indicator for zero debt	0.658	0.346		-0.415	-0.110	
Log (1 + salary of the respondent)	0.140	2.676	***	0.110	1.146	
Log (1 + salary of the spouse)	0.093	2.517	**	0.180	2.543	**
Indicator for the desire to leave a bequest	0.180	0.436		1.797	2.259	**
Indicator for foreseeable major financial obligation	0.563	1.478		1.422	1.935	*
Log (1+ sizable inheritance expected)	-0.116	-0.518		-0.147	-0.344	
Indicator for zero inheritance expected	-1.807	-0.632		-2.387	-0.434	
Average age of the couple	0.340	2.102	**	-0.361	-1.089	
Squared average age of the couple	-0.004	-1.926	*	0.007	1.850	*
Education level of the resondent	0.345	3.312	***	-0.105	-0.509	
Education level of the spouse	0.197	2.035	**	0.234	1.214	
Financial Vulnerability Index (IMPACT)	1.314	3.591		1.010	1.516	
Indicator for IMPACT $\geqslant$	-3.038	-2.127	**	-2.330	-0.912	
Log(1+Net amount at risk of whole life insurance)	-0.270	-8.267	***			
Log(1+Face value of term life insurance)				-0.452	-7.600	***

Notes: Number of observations: 2,150.

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level

4 Concluding Remarks

This paper explores a multivariate two-part framework for household ownership of life insurance. We use statistical models of multivariate binary data and multivariate severity data. When joint insurance ownership or claims behavior are analyzed, these tools can be of significant use. In our case, they help to improve our understanding of a household's life insurance demand.

We find that variables such as the amount of stock, bond, real estate and debt a household owns and income earned can have a significant impact on life insurance demand. However, the impact on the frequency and severity part as well as the type of life insurance can differ even in terms of the sign of coefficients. For example, the greater the bond holdings of a household, the less likely it is to own life insurance. Further, if the household owns life insurance, the greater the bond holdings, the more life insurance it tends to own. Some demographic characteristics, such as age and education level, may affect life insurance demand. In general, older and better educated households have a higher demand for life insurance. Also, the financial vulnerability index proposed by Lin and Grace (2007) proves to be a significant variable in explaining households life insurance demand.

One important contribution of our study is that we find that the demand of term and whole life insurance are substitutes in frequency and complements in severity after controlling for all the explanatory variables. This mixed effect extends prior work which established that term life insurance and whole life insurance are substitutes for one another. As a robustness check, we also examined life insurance demand for single person households in Appendix D. The results for this subpopulation supports our findings for the married couples subpopulation.

This paper provides a better understanding of factors that drive life insurance demand; this can help life insurers target markets more effectively. In the future, we hope to use the probability sampling basis of these data to extrapolate our findings to the national population; such results would be useful for public policy purposes.

References

Bernheim, B.D., K.G. Carman, J. Gokhale, and L.J. Kotlikoff, 2001. The Mismatch Between Life Insurance Holdings and Financial Vulnerabilities: Evidence From the Survey of Consumer Finances. NBER working paper no. 8544.

Burnett, J.J., and B.A. Palmer, 1994, Examining Life Insurance Ownership Through Demographic and Psychographic Characteristics. *Journal of Risk and Insurance* Vol. 51, No. 3, pp. 453-467.

Chavas, J.P., 2004. *Risk Analysis in Theory and Practice*. Elsevier Academic Press, San Diego, California.

Cragg, J.G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 39(5), pp. 829-844.

Frees, E.W., 2010. *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press, New York.

Frees, E.W. and E.A. Valdez, 1998. Understanding Relationships Using Copulas. *North American Actuarial Journal*, Vol. 2, pp. 1- 25.

Frees, E.W. and P. Wang, 2005. Credibility Using Copulas. *North American Actuarial Journal*, Vol. 9, pp. 31-48.

Frees, E.W., J. Gao, and M.A. Rosenberg, 2009. Predicting the Frequency and Amount of Health Care Expenditures. Working paper.

Genest, C. and A.-C. Favre, 2007. Everything you always wanted to know about Copula Modeling but were afraid to ask. *Journal of Hydrologic Engineering* 12, 347-368.

Goldsmith A. 1983. Household Life Cycle Protection: Human Capital versus Life Insurance. *Journal of Risk and Insurance* Vol.50, No.3., pp. 473-486.

Greene, W. 2008. *Econometric Analysis, Sixth Edition*. Pearson Prentice-Hall, Upper Saddle River, New Jersey.

Halek. M and J. G. Eisenhauer, 2001. Demography of Risk Aversion. *Journal of Risk and Insurance* Vol.68, No.1, pp. 1-24.

Insurance Information Institute FACTS AND STATISTICS. *Insurance Information Institute* website: http://www.iii.org/media/facts/statsbyissue/life/

Kleiber, C. and S. Kotz, 2003. *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, New York

Klugman, S.A., H.H. Panjer and G. E. Willmot (2008). Loss Models: From Data to Decisions. John Wiley & Sons, Hoboken, New Jersey.

Lin, Y, and M.F. Grace, 2007. Household Life Cycle Protection: Life Insurance Holdings, Financial Vulnerability, and Portfolio Implications. *Journal of Risk and Insurance* Vol.74, No.1, pp.141-173.

McCullagh, P. and J.A. Nelder, 1989. Generalized Linear Models, 2nd ed. Chapman & Hall, London.

Nelson, R.B. 1999. *An Introduction to Copulas*, Lecture Notes in Statistics 139, Springer.

The Federal Reserve Board, 2004. Codebook for the 2004 Survey of Consumer Finances

Tobin, J (1958). Estimation of relationships for limited dependent variables. *Econometrica* 26, pp. 24-36.

Vidal-Meliá, C. and A. Lejárraga-García (2005). The Bequest Motive and Single People's Demand for Life Annuities. *The Belgian Actuarial Bulletin*, 4, pp. 4-15.

Zeger, C.V. and P. Diggle, 1992. Modeling Multivariate Binary Data with Alternating Logistic Regressions. *Biometrika*, Vol. 80, No.3, pp. 517-526.

Zellner, A. and T.H. Lee, 1965. Joint Estimation of Relationships Involving Discrete Random Variables. *Econometrica*, Vol. 33, pp. 382-394.

A Appendix: Financial Vulnerability Index

This section describes the financial vulnerability index defined by Lin and Grace (2007). For the calculation, the SCF variables needed are:

1. Household total labor and nonlabor income before taxes and deductions, $Tincome_i$
2. Salary of the respondent ($y_{res,i}$) and spouse ($y_{spouse,i}$).
3. Age of respondent (x) and spouse (y).
4. Number of children N

In addition to the SCF variables, the calculation requires mortality and consumption to income information. We used the mortality data base available from the Society of Actuaries' that can be found at `http://www.soa.org/professional-interests/technology/tech-table-manager.aspx`. The consumption to income information is available from the US Bureau of Labor Statistics at `http://www.bls.gov/cex/csxann04.pdf`.

The financial vulnerability index (IMPACT_i) measures the financial impact (living standard decline) of the death of one member of the household on the rest. It is calculated according the following steps.

1. The living standard (C_i) is the per person consumption level. It is based on the total income of the household and the population consumption-to-income ratio for the household income bracket, adjusted for number of household members and household scale economies,

$$C_i = \alpha_i \frac{Tincome_i}{(2 + \frac{N}{2})^{0.678}}.$$

   Here, α_i is the consumption-to-income ratio for the general population in the $Tincome_i$ income bracket. The "2" indicates there are two adults in the household and 0.678 measures household scale economies. This suggests that in order to achieve the same living standard, a two-adult household must spend 1.6 times ($2^{0.678}$) as much as an one-adult household. (Bernheim et al. (2001))

2. When the respondent dies, the reduced living standard of the spouse becomes

$$C_{spouse,i} = \beta_{spouse,i} \frac{Tincome_i - y_{res,i}}{(1 + \frac{N}{2})^{0.678}}.$$

   Here, $\beta_{spouse,i}$ is the spouse's consumption-to-income ratio if the respondent dies.

3. The relative impact on the household if the respondent dies can be expressed as a percentage decline

$$\mathrm{IMPACT}_{spouse,i} = \frac{C_{spouse,i}}{C_i} - 1 = \frac{\beta_{spouse,i}(\mathrm{Tincome}_i - y_{res,i})(2 + \frac{N}{2})^{0.678}}{\alpha_i \mathrm{Tincome}_i (1 + \frac{N}{2})^{0.678}} - 1.$$

4. The absolute impact on the household if respondent dies is

$$\text{IMPACT}_{spouse,i} * y_{res,i}$$

5. If the respondent dies at age x, the household will incur annual absolute living standard decline for (65-x) years. The age effect of the death of the respondent at age x can be captured by an annuity factor $a_{\overline{65-x}|}$. This is an annuity certain.

6. The current life insurance holding reflects the household's expectation of its potential risks if one of the spouses dies in the foreseeable future, e.g., one year. The one-year death probabilities $q_{x,i}^{res}$ for the respondent reflects such concern.

7. The impact on the household if the spouse dies can be obtained by reversing *res* and *spouse*.

8. Taking into account all of the above factors, the index of financial vulnerability (IMPACT_i) of the household i is defined in a way similar to the definition of standard deviation,

   IMPACT_i=

$$\sqrt{q_{x,i}^{res}(\text{IMPACT}_{spouse,i} \cdot y_{res,i} \cdot a_{\overline{65-x}|})^2 + q_{y,i}^{spouse}(\text{IMPACT}_{res,i} \cdot y_{spouse,i} \cdot a_{\overline{65-y}|})^2}.$$

B Appendix: Variables

Variable	Description
Dependent Variable	
Term	Indicator for term life insurance demand (=1 own term life insurance)
WHOLE	Indicator for whole life insurance demand (=1 own whole life insurance)
INSURANCE	Indicator of Insurance Purchase (=0 no insurance, =1 term life insurance only, =2 whole life insurance only, =3 Both term and whole life insurance)
FACETerm	Face value of term life insurance
NAR	Whole life insurance net amount at risk
Independent Variable	
CASHEQV	Cash and cash equivalent
INDCASHEQV	Indicator for zero cash and cash equivalent (=1 zero cash)
FUND	Mutual fund
INDFUND	Indicator for zero fund (=1 zero fund)
STOCK	Stock
INDSTOCK	Indicator for zero stock (=1 zero stock)
BOND	Bond
INDBOND	Indicator for zero bond (=1 zero bond)
ANNUITY	Individual annuity not including job pension
INDANNUITY	Indicator for zero annuity (=1 zero annuity)
RETIREMENT	A households individual retirement account
INDRETIREMENT	Indicator for zero retirement (=1 zero retirement account)
REALESTATE	Real estate
INDREALESTATE	Indicator for zero real estate (=1 zero real estate)
OTHASSETS	Other assets
INDOTHASSETS	Indicator for zero other assets (=1 zero other assets)
DEBT	Total debt of the household
INDDEBT	Indicator for zero debt (=1 zero debt)
SALARY1	Salary and wages of respondent before taxes
SALARY2	Salary and wages of spouse before taxes
INCOME	A households total income before taxes
BEQUEST	Desire to leave a bequest (=1 Yes)
OBLIGATION	Foreseeable major financial obligations (=1 Yes)
INHERITANCEExp	Sizable inheritance expected
INDINHERITANCE	Indicator for zero inheritance expected (=1 zero inheritance expected)
AGE	Average age of couple
EDUCATION1	Education of respondent
EDUCATION2	Education of spouse/partner
IMPACT	Financial vulnerability index
INDIMPACT	Indicator for IMPACT $\geqslant 4$(=1 if IMPACT$\geqslant$ 4)

C Appendix: Robustness Checks

C.1 Frequency Component

In order to check the robustness of our result, we fit an alternative multivariate frequency models to our data, the multinomial logit regression model.

A multinomial logit regression model is a natural extension of the logistic regression model (e.g., Frees (2010)). The multinomial logit regression model assumes the various categories of the dependent variable response differently to the same set of explanatory variables. When we use multinomial logit regression, we choose one category of the dependent variable as the reference category. We can represent the linear combination of explanatory variables as the log odds ratio of a certain category of the dependent variables relative to the reference category.

The variable of interest is r_i, which can take values $1, 2, ..., c$ (c mutually exclusive categories in total). We choose c as the reference category, then for a single observation r_i,

$$\ln \frac{\Pr(r_i = j)}{\Pr(r_i = c)} = \mathbf{x}_i' \boldsymbol{\beta}_j, \quad j = 1, \ldots, c-1.$$

For each category of r_i,

$$\begin{aligned} \Pr(r_i = j) &= \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta}_j)}{\sum_{j=1}^{c} \exp(\mathbf{x}_i' \boldsymbol{\beta}_j)}, \quad j = 1, \ldots, c-1 \\ \Pr(r_i = c) &= \frac{1}{\sum_{j=1}^{c} \exp(\mathbf{x}_i' \boldsymbol{\beta}_j)}. \end{aligned}$$

With this, the log-likelihood of the ith observation is

$$l_i = \mathbf{1}(r_i = c) \ln \frac{1}{\sum_{j=1}^{c} \mathbf{x}_i' \boldsymbol{\beta}_j} + \sum_{j=1}^{c-1} \mathbf{1}(r_i = j) \ln \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta}_j)}{\sum_{j=1}^{c} \mathbf{x}_i' \boldsymbol{\beta}_j}$$

where $\mathbf{1}(\cdot)$ is an indicator function.

We can express the likelihood of all observations as $\sum_{i=1}^{n} l_i$. By maximizing the log-likelihood, we can obtain reliable parameter estimates.

For our application, the multinomial logit regression model has a four-level dependent variable INSURANCE: no life insurance purchase (INSURANCE=0), term life insurance only (INSURANCE=1), whole life insurance only (INSURANCE=2) and both term life insurance and whole life insurance (INSURANCE=3). We choose no life insurance purchase as the reference level and all the covariates estimated for level i are based on a comparison between level i and the reference level.

Using multinomial logit regression in our case means estimating a system of three equations simultaneously. There are 31 covariates for each level and therefore 93 coefficients are estimated. Compared with single equation estimation, the joint estimation of a system of equations is more efficient (e.g., Zellner and Lee (1965)). Though the dependent variables are not exactly the same, Table 7 still reveals a similar results as Table 4 for the marginal distribution. Given a household's characteristics, one can examine the association between these two types of life insurance using multinomial logit regression model.

Table 7: Multinomial Logit Regression

	Term Life Only (992)			Whole Life Only (294)			Both Term and Whole (424)		
Parameter	Estimate	t-statistic		Estimate	t-statistic		Estimate	t-statistic	
Intercept	2.484	1.152		0.643	0.245		0.199	0.077	
Log (1+ cash and cash equivalent)	0.033	0.806		0.043	0.831		0.144	2.842	***
Indicator for Izero cash and cash equivalent	-0.361	-0.804		0.461	0.731		-0.222	-0.199	
Log (1+ fund)	0.090	1.504		-0.007	-0.106		-0.013	-0.202	
Indicator for zero fund	0.894	1.235		-0.458	-0.556		-0.593	-0.751	
Log (1+stock)	-0.124	-2.545	**	-0.094	-1.680	*	-0.182	-3.414	***
Indicator for zero stock	-0.932	-1.728	*	-0.834	-1.291		-1.803	-3.009	***
Log (1+ bond)	-0.138	-3.443	***	-0.144	-3.210	***	-0.170	-3.933	***
Indicator for zero bond	-1.394	-3.399	***	-1.638	-3.407	***	-2.165	-4.838	***
Log (1+ annuity)	-0.180	-1.689	*	-0.097	-0.893		-0.152	-1.349	
Indicator for zero annuity	-2.793	-2.033	**	-2.219	-1.546		-2.579	-1.765	*
Log (1+ retirement)	0.099	1.924	*	0.019	0.304		0.011	0.187	
Indicator for zero retirement	0.044	0.080		-0.652	-0.931		-0.628	-0.936	
Log (1+ real estate)	-0.394	-4.609	***	-0.009	-0.086		-0.246	-2.380	**
Indicator for zero real estate	-4.836	-4.926	***	-0.613	-0.497		-3.575	-2.911	***
Log (1+ other assets)	0.089	1.474		0.065	0.926		0.106	1.539	
Indicator for zero other assets	0.769	1.145		-0.280	-0.318		0.195	0.230	
Log (1 + debt)	0.088	1.724	*	0.021	0.316		0.113	1.721	
Indicator for zero debt	0.359	0.600		0.165	0.207		0.302	0.370	
Log (1 + salary of the respondent)	0.053	3.064	***	0.041	1.976	*	0.044	2.172	**
Log (1 + salary of the spouse)	0.029	2.189	**	0.028	1.705	*	0.058	3.696	***
Indicator for the desire to leave a bequest	0.274	1.819	*	0.612	3.175	***	0.259	1.450	
Indicator for foreseeable major financial obligation	0.148	1.098		0.035	0.206		0.144	0.887	
Log (1+ sizable inheritance expected)	-0.108	-1.343		-0.120	-1.197		-0.083	-0.882	
Indicator for zero inheritance expected	-1.347	-1.303		-1.467	-1.125		-1.171	-0.963	
Average age of the couple	0.137	2.649	**	0.086	1.174		0.143	1.943	*
Squared average age of the couple	-0.001	-2.479	**	-0.001	-0.774		-0.001	-1.277	
Education level of the resondent	0.096	2.828	***	-0.015	-0.338		0.080	1.787	*
Education level of the spouse	0.061	2.010	**	0.076	1.772	*	0.065	1.560	
Financial Vulnerability Index (IMPACT)	0.401	2.393	**	0.238	1.280		0.500	2.826	***
Indicator for IMPACT $\geqslant$4	-1.413	-2.211	**	-1.067	-1.510		-1.498	-2.238	**

Notes: Number of observations: 2,150. The reference level is No life insurance purchase (440).

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level

C.2 Severity Component

The robustness for our severity part result is checked by mimicking a bivariate lognormal distribution. The logged face value of term life insurance and the logged NAR of whole life insurance are fitted by two OLS respectively. Then we calculated the Spearman correlation for the residuals from each model, for those observations owning both types of life insurance. The Spearman correlation is 0.09925. This results confirms that the relationship between the face value of term life insurance and the NAR of whole life insurance is positive.

D Appendix: Single Person Households

Single person households consist of individuals who are single (never married, separated, divorced or widowed). For this kind of household, the respondent is the policyholder as well as the decision-maker in terms of life insurance demand. Therefore, there is no ambiguity in assessing an individual's preferences.

In order to compare with the result of married couple households, we focus on the subsample of single individuals with the age from 20 to 64 and exclude those observations where labor income is paid irregularly. The number of observations for single person households is 643, among which 55% own at least one type of life insurance. We include all the explanatory variables from the married couple analysis except the financial vulnerability index and those variables related to the spouse of the respondent, which is unavailable for single person households. Table 8 displays the result for the frequency part. We can see that most coefficients are of the same sign as in the married couple case. In particular, the correlation variable Rho is also negative and significant.

Table 8: Bivariate Probit Regression for Single Person Households

	Term Life			Whole Life		
Parameter	Estimate	t-statistic		Estimate	t-statistic	
Intercept	1.882	1.099		-1.594	-0.808	
Log (1+ cash and cash equivalent)	0.045	1.434		0.010	0.250	
Indicator for zero cash and cash equivalent	0.114	0.402		0.084	0.231	
Log (1+ fund)	-0.071	-1.116		0.032	0.423	
Indicator for zero fund	-0.566	-0.843		0.608	0.746	
Log (1+stock)	-0.116	-1.918	*	-0.027	-0.411	
Indicator for zero stock	-1.109	-1.893	*	-0.084	-0.129	
Log (1+ bond)	-0.050	-0.808		-0.121	-1.718	*
Indicator for zero bond	-0.329	-0.640		-0.862	-1.495	
Log (1+ annuity)	-0.064	-0.640		-0.146	-1.375	
Indicator for zero annuity	-0.710	-0.641		-1.437	-1.248	
Log (1+ retirement)	0.110	2.259	**	0.072	1.296	
Indicator for zero retirement	0.797	1.647	*	0.505	0.886	
Log (1+ real estate)	-0.166	-2.194	**	-0.010	-0.117	
Indicator for zero real estate	-1.871	-2.200	**	-0.406	-0.422	
Log (1+ other assets)	-0.053	-0.974		0.081	1.294	
Indicator for zero other assets	-0.629	-1.226		0.499	0.826	
Log (1 + debt)	0.051	1.281		-0.024	-0.532	
Indicator for zero debt	0.232	0.593		-0.439	-0.932	
Log (1+ salary of the respondent)	0.069	4.966	***	-0.002	-0.146	
Indicator for the desire to leave a bequest	0.192	1.521		0.148	1.002	
Indicator for foreseeable major financial obligation	0.062	0.559		0.480	3.538	***
Log (1+ sizable inheritance expected)	-0.072	-0.957		0.032	0.363	
Indicator for zero inheritance expected	-0.432	-0.491		0.238	0.226	
Age of the respondent	0.044	1.297		-0.004	-0.088	
Squared age of the respondent	0.000	-0.904		0.000	0.506	
Education level of the respondent	0.010	0.403		0.045	1.365	
Rho	-0.325	-4.219	***			

Notes: Number of observations: 643

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level

Table 9 displays the result from the severity model, which is based on the subsample of single person households that own life insurance (356 observations). Again, just a few variables have signs that differ from the married couple case. The correlation variable in the gaussian copula, Rho, is positive though insignificant. This weakly confirms that after controlling for all the explanatory variables, the amount of two types of life insurance purchase by a household is complements to each other.

Table 9: Gaussian Copula with Gamma Marginals for Single Person Households

	Term Life			Whole Life		
Parameter	Estimate	t-statistic		Estimate	t-statistic	
Intercept	-1.175	-0.629		-9.387	-2.322	
Log (1+ cash and cash equivalent)	0.007	0.187		-0.227	-2.302	**
Indicator for Izero cash and cash equivalent	-0.240	-0.672		-2.014	-2.253	**
Log (1+ fund)	0.141	1.791		0.416	3.083	***
Indicator for zero fund	1.737	2.170	**	4.028	2.642	**
Log (1+stock)	0.026	0.353		-0.294	-2.240	**
Indicator for zero stock	0.180	0.266		-2.337	-1.705	*
Log (1+ bond)	0.204	2.627	***	0.671	3.400	***
Indicator for zero bond	1.667	2.763	***	4.707	3.387	***
Log (1+ annuity)	-0.114	-1.049		-0.147	-0.872	
Indicator for zero annuity	-0.294	-0.253		-0.126	-0.070	
Log (1+ retirement)	0.135	2.445	**	0.126	1.144	
Indicator for zero retirement	1.365	2.425	**	1.650	1.385	
Log (1+ real estate)	0.291	3.021	***	0.661	4.014	***
Indicator for zero real estate	3.230	3.021	***	7.390	3.854	***
Log (1+ other assets)	0.173	2.615	**	0.073	0.668	
Indicator for zero other assets	1.664	2.636	**	0.385	0.339	
Log (1 + debt)	-0.007	-0.159		0.083	0.928	
Indicator for zero debt	-0.275	-0.630		0.728	0.778	
Log (1+ salary of the respondent)	0.087	4.776	***	0.019	0.592	
Indicator for the desire to leave a bequest	0.154	1.148		0.249	0.808	
Indicator for foreseeable major financial obligation	-0.040	-0.314		0.365	1.199	
Log (1+ sizable inheritance expected)	-0.057	-0.510		0.141	0.824	
Indicator for zero inheritance expected	-0.748	-0.566		1.235	0.586	
Age of the respondent	0.097	6.498	***	0.127	4.863	***
Squared age of the respondent	-0.001	-8.250	***	-0.002	-6.923	***
Education level of the resondent	0.107	3.580	***	0.146	1.705	*
Scale	1.066	0.080	$	0.880	0.081	$
Rho	0.142	0.719				

Notes: Number of observations: 356.

$ This is the standard error for the scale parameter.

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level